2000-12

Feasibility study of speech recognition technologies for operating within a medical First Responder's environment

Harris, Leroy W.

Monterey, California. Naval Postgraduate School

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

FEASIBILITY STUDY OF SPEECH RECOGNITION TECHNOLOGIES FOR OPERATING WITHIN A MEDICAL FIRST RESPONDER'S ENVIRONMENT

by

Leroy W. Harris Jr.

December 2000

Thesis Co-Advisors:	Monigue P. Fargues
	Ray T. Clifford
Associate Advisor:	Douglas E. Brinkley

Approved for public release; distribution is unlimited.

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** December 2000	**3. REPORT TYPE AND DATES COVERED** Master's Thesis

4. TITLE AND SUBTITLE : Feasibility Study Of Speech Recognition Technologies For Operating Within A Medical First Responder's Environment	**5. FUNDING NUMBERS**
6. AUTHOR(S) Harris, Leroy W. Jr.	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A	**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**

11. SUPPLEMENTARY NOTES

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.	**12b. DISTRIBUTION CODE**

13. ABSTRACT ***(maximum 200 words)***

This thesis was designed to address some of the issues facing the medical First Responder who is continually tasked with providing care within multi-national environments. Currently, there are no established billets or quota requirements at the Defense Language Institute Foreign Language Center for Navy Corpsmen for the purposes of foreign language education prior to an overseas assignment or deployment.

The primary Speech Recognition (SR) device used in this study was the Voice Response Translator (VRT). Navy Corpsmen and Army Medics were asked to evaluate the VRT's capabilities in assisting with non-English speaking patient assessments. Other SR assisted technologies available to overcome some of the burden of providing healthcare in a foreign language environment were also studied. The results of this feasibility study show that SR assisted technologies are a viable tool available for operation within a medical First Responder's environment.

14. SUBJECT TERMS Speech Recognition, Machine Translation, Field Medicine, Medical Support, Medical.			**15. NUMBER OF PAGES**
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

Approved for public release; distribution is unlimited

FEASIBILITY STUDY OF SPEECH RECOGNITION TECHNOLOGIES FOR OPERATING WITHIN A MEDICAL FIRST RESPONDER'S ENVIRONMENT

Leroy W. Harris Jr.
Lieutenant, U.S. Navy
B.A., Saint Leo College, 1993

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN INFORMATION SYSTEMS TECHNOLOGY

from the

NAVAL POSTGRADUATE SCHOOL
December 2000

Author: [redacted]
Leroy W. Harris Jr.

Approved by: [redacted]
Monigue P. Fargues, Co-Thesis Advisor

[redacted]
Ray T. Clifford, Co-Thesis Advisor

[redacted]
Douglas E. Brinkley, Associate Advisor

[redacted]
Dan Boger, Chairman
Information Systems Academic Group

ABSTRACT

This thesis was designed to address some of the issues facing the medical First Responder who is continually tasked with providing care within multi-national environments. Currently, there are no established billets or quota requirements at the Defense Language Institute Foreign Language Center for Navy Corpsmen for the purposes of foreign language education prior to an overseas assignment or deployment.

The primary Speech Recognition (SR) device used in this study was the Voice Response Translator (VRT). Navy Corpsmen and Army Medics were asked to evaluate the VRT's capabilities in assisting with non-English speaking patient assessments. Other SR assisted technologies available to overcome some of the burden of providing healthcare in a foreign language environment were also studied. The results of this feasibility study show that SR assisted technologies are a viable tool available for operation within a medical First Responder's environment.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENT

First and foremost, I acknowledge my beloved family Soana, Natasha, and Benjamin. Without your continued encouragement and support, I would not have achieved our past successes nor would I have reached this milestone in our lives.

I wish to express my sincere appreciation to Dr Ray Clifford, Professors Monique Fargues and Douglas Brinkley, and Lt Col Steve Butler for their untiring leadership, motivation and speedy turnaround time which was the driving force that maintained the research focus and matured this thesis to completion.

I acknowledge and thank Mr. Tim McCune, Lieutenant Commanders Kurt Henry and Eric Rasmussen, Mrs. Christine Montgomery and Lieutenant John Kendrick who provided their support throughout the information gathering phase of this thesis.

I. INTRODUCTION

A. INTRODUCTION

The Navy Medical Department's mission is "Support the combat readiness of the uniformed services and to promote, protect and maintain the health of all those entrusted to our care, anytime, anywhere." In addition, the primary focus of its goal under Force Health Protection is "The medical departments must be prepared to respond effectively and rapidly to the entire spectrum of potential military operations –from major regional contingencies to Military Operations Other Than War (MOOTW). Readiness to support wartime/contingency operations will require us to successfully accomplish several missions simultaneously. We must be able to identify the medical threat; develop medical organizations and systems to support potential combat scenarios; train medical units and personnel for their wartime roles. We must train non-medical personnel in medical subjects; conduct medical research to discover new techniques and materiel to conserve fighting strength; and provide both preventive and restorative health care to the military force." [Ref .1]

This thesis was designed to address some of the issues facing the medical First Responder who is continually tasked with providing care within multi-national environments. Little consideration is given to language training prior to preparation and during the operation for the foreign language barriers to be encountered while carrying out their mission. Humanitarian and Peace Keeping operations continue to rise, while

medical personnel assets continue to decline. Currently, there are no established billets or quota requirements at the Defense Language Institute Foreign Language Center for Navy Medical Department personnel for the purposes of foreign language education prior to an overseas assignment or deployment. The Navy Medical Department is very diverse and has many individuals who speak and/or understand a foreign language. Medical personnel, however, are normally assigned to duty according to their professional expertise and the needs of the Navy and Marine Corps communities they serve and not solely based on their language proficiency. Any incident where there is someone fluent in the native language, of the area where language is a barrier, would be considered a unique and special occasion.

There seems to be an assumption that providing healthcare support in foreign environments is universal and that medical personnel will carry out their healthcare delivery mission regardless of the language barriers encountered. The Navy Medical Department has proven its ability to deliver healthcare in a multinational environment, however there are Speech Recognition (SR) and other assisted technologies available to overcome some of the burden of providing healthcare in a foreign language environment. This feasibility study will outline SR assisted technologies available for operation within a medical First Responder's environment.

B. PURPOSE OF THE STUDY

The principal objective of this study is to identify SR devices that can be deployed in a medical first responder's operating environment where language is considered a barrier.

C. RESEARCH QUESTIONS

The literature review, research questionnaire, demonstration and evaluation for this thesis were designed to collect data to address the following proposed research questions:

1. What are the SR technologies available for operating within a medical first responder's environment?

2. What are the SR technologies currently being used in a medical first responder's environment?

3. What are the SR technologies available for operating within a medical first responder's foreign language environment?

4. What are the SR technologies currently being used in a medical first responder's foreign language environment?

5. What other SR assisted technologies are being used in other than a medical environment could be feasible for operation within a medical first responder's environment?

D. SCOPE OF THE STUDY

The scope of this study includes a review of the history of computer speech technology and research of SR devices available for operating within a medical first responder's environment. The study also conducted a demonstration and evaluation of SR devices using a foreign language within a simulated medical environment. Finally, the study concludes with a recommendation to the Navy Medical Department concerning SR devices to be considered in future field demonstrations and evaluations.

E. METHODOLOGY

The methodology used in this study consisted of an in-depth analysis and evaluation of SR technologies and devices through a literature review, consultations with computer speech technology and foreign language experts, and a practical demonstration and evaluation of the Voice Response Translator (VRT) and the Multi-lingual Interview System (MIS). The literature review consisted of:

- Internet search of SR subjects on websites and homepages (DoD, academic, and commercial)
- A MEDLINE Literature index search of SR subjects through the National Library of Medicine.
- A Computer select database search of SR subject at the Naval Postgraduate School Library.
- Review of various studies, reports and other documentation related to SR projects and issues, both within the DoD and the private sector.

The consultation efforts consisted of:

- Attendance at the June 1999 Language Workshop Meeting, Office of Special Technology (OST).
- Attendance at the 1999 Healthcare Information and Management Systems Society Conference.
- Attendance at the 1999 American College of Healthcare Executives Conference.
- Collaboration with Naval Aerospace Medical Research Laboratory project officers on the MIS.
- Collaboration with the Integrated Wave Technologies, Inc. project officer on the VRT.

- Collaboration with U.S. Army Research Laboratory project officers on the FALCon.

- Collaboration with the Language Systems, Inc. project officer on the Voice-to-Voice Language Translation.

- Collaboration with the Defense Language Institute Foreign Language Center.

The demonstration and evaluation consist of:

- Developing a SR demonstration and evaluation questionnaire instrument.

- Demonstrating the use of the VRT.

- Demonstrating the use of the MIS.

F. THESIS ORGANIZATION

This thesis is composed of six chapters. This chapter provides the introduction, purpose of the study, research questions, scope and methodology employed to conduct the research. Chapter II provides an historical view of computer speech technology. Chapter III describes some current SR initiatives in the DoD and private sector. Chapter IV describes the methodology for the demonstration and evaluation of the VRT and MIS. Chapter V discusses the demonstration and evaluation results. Chapter VI provides the conclusion and recommendations for future research.

G. BENEFITS OF THE STUDY

This thesis provides a reference of ongoing SR technologies being developed that can be applied in a medical first responder's operating environment. These results will be used to propose SR technologies to the Department of the Navy Bureau of Medicine and

Surgery for consideration in humanitarian, peacekeeping, deployable and overseas environments.

II. OVERVIEW OF COMPUTER SPEECH TECHNOLOGY

A. THE ERA OF ARPA

In 1971, the Advanced Projects Research Agency (ARPA) challenged American companies and universities to develop a speech-understanding system with a vocabulary of at least 1,000 words capable of processing connected speech with an error rate less than ten percent in a low-noise environment for use by many cooperative speakers. The systems were allowed to have an artificial syntax and a highly constrained context and were not required to operate in real time, as discussed in [Ref. 2]. ARPA deliberately used the word understanding, as opposed to recognition. Understanding, when used in this way, came to mean that once input was recognized, or partially recognized, it would be further processed. If a question was posed, the system would be required to answer it; if a request was made, the system would have to fulfill it, as discussed in [Ref. 3].

At the end of the project in late 1976, three contractors, Carnegie Mellon University (CMU), Bolt Beranek and Newman (BBN), and System Development Corporation (SDC) - Stanford Research Institute (SRI), had produced six systems. The three most viable were the *Harpy* and *Hearsay II* systems of CMU and the *Hwim* ("Hear what I mean") system of BBN. Of these, only *Harpy* fully met the five-year goals of ARPA. Details of these and other ARPA project systems may be found in [Ref. 2]. The ARPA project pioneered the use of linguistic knowledge. *Hearsay II* borrowed the "blackboard" notion from the artificial intelligence field. Blackboard is jargon for a database of information made

available to the diverse processes of a software system. *Hearsay II* had various subparts that checked on whether a potential sound sequence was consistent with syllable structure, whether a potential syllable combination was a legitimate word, whether a potential word combination was a legitimate phrase, and so on. Through the blackboard, information from these various levels of knowledge sources could be exchanged. Thus, if a potential word was found in *Hearsay II's* dictionary of allowable words, the system could back up and substitute a different sound or syllable, forming a different word, which it could then try out. *Hwin* employed a syntactic analyzer called an "augmented transition network" that eliminated phonetic choices that led to ungrammatical sentences. *Harpy* achieved a similar end by means of a "finite state grammar" (Both of these syntactic analyzers are described in [Ref. 4]). In both systems, the syntactic component would ask the recognizer for its next best guess and continue to do so until a grammatically acceptable sequence occurred when the recognizer's best guess was ill-formed, such as, "John green its dog." The system rejected the input as unrecognizable if no well-formed sentence could be found. All large speech recognition systems developed after ARPA had ways to restrict recognition choices based on the syntactic constraints of the language as discussed in [Ref. 3].

1. Noise Consideration

The ARPA projects were concerned chiefly with the kinds of fundamental problems of recognition and understanding, but none worried about noise. Experiments took place in quiet environments using high-quality electronics. The quest for practical, usable

systems led to an investigation of the effects of noise, which can be devastating. Systems with five percent error rates in quiet environments found themselves with 35 percent error rates when background noise was introduced. Channel noise plays havoc with the recognition process as does noise introduced by the speaker such as coughing, throat clearing, snuffling, snorting, sputtering, spluttering, stuttering, stammering, slurring, lisping, lip smacking, and nonlinguistic vocalizations such as hemming, hawing, uh-ing, and er-ing. These difficulties were addressed throughout the 1980s. Advances in electronics led to improved noise-canceling microphones. An understanding of the distortions introduced by the telephone network allowed them to be modeled and accounted for during the recognition process. Some extraneous sounds introduced by speakers could be detected and ignored during recognition. Human factors experts addressed the problem of getting users to speak fluently. In all, immunity to noise improved greatly and led to widespread applications, such as voice dialing a mobile phone in a moving automobile, etc. as discussed in [Ref. 3].

2. Cost Consideration

The ARPA project speech recognition systems would be extremely expensive if they were available for sale in the commercial markets. The post-ARPA history of speech recognition saw the price tumble, much as it did for desktop computers. Speech recognition systems are classified and priced a little like automobiles. Your basic car has a stick shift and an AM radio and no air conditioning or power windows. Your basic speech recognizer only accepts speech spoken with pauses between each word, must be

pretrained to your voice, and is limited in vocabulary. A few more dollars will get you an automatic transmission or stereo system in your new car. Likewise, spending some extra money will get you a speech recognizer that lets you speak continuously without pausing between words, and may recognize the speech of your friends if their voices are similar to yours, as discussed in [Ref. 3].

3. Abbreviations

Written English uses thousands of abbreviations and many are standard in their use. Table 1 below lists abbreviations and an indication of how they might be pronounced.

Abbreviation	**Spoken As**
Mrs.	Missus
Dr.	Doctor, Drive
Ph.D.	Pee Aitch Dee
St.	Street, Saint, Stanza
Ch.	Chapter, Chaplain
5:30 A.M.	Five Thirty A EM

Table 1: Abbreviations Compared To Spoken Words.

Common abbreviations may be put in a dictionary. Ambiguities are resolved by context or frequency of occurrence. Using context, it is easy to disambiguate that Dr Einstein lives on Riverside Dr., and using statistics one would choose to pronounce the abbreviation *Ch.* as *chapter*, that being a more common usage than *chaplain*. Of course, Ch is most likely to abbreviate *Chapter* when it is followed by a numeral of any kind. A

good text-to-speech program handles numbers (both cardinal and ordinal), fractional expressions, decimal numbers, dates, and times of day, currency amounts, and punctuation. The period and comma are represented by pauses of varying lengths. The colon and semicolon engender somewhat shorter pauses. The question mark produces rising intonation, as discussed in [Ref. 3].

4. Microphone Consideration

Inside every microphone is a diaphragm, capable of vibrating in concert with any sound whose frequencies are within its range of operation. These oscillations are converted into electrical signals in a variety of ways depending on the type of microphone. In a carbon microphone, often found in telephones, the level of resistance in an electrical circuit is controlled by the oscillations so that a variation in electrical output replicates the original sound. A variation in capacitance produces the same effect in a condenser microphone. Vibration induced variations of electromagnetic fields and shapes of piezoelectric crystals are also used to control the transducing of sound to an electrical signal. Microphones are designed for various patterns of reception and various placements in the environment. Omni-directional microphones collect sounds from all directions. Uni- and bi-directional microphones have maximum sensitivity to sound coming from one or two directions. Microphones may be handheld, the favorite of rock singers; attached to the lapel, the favorite of talk show guests; head-worn, the favorite of telephone operators; hung from tall ceilings, the favorite of concert pianists; or stuck in the ear, nobody's favorite. Noise-canceling microphones are important when noise is not

well tolerated, as in computer speech recognition. A typical noise-canceling microphone is actually two microphones, one directed at the speaker and the other in the opposite direction. Ambient noise enters both microphones at about equal levels of amplitude, but the amplitude level of speech is much higher in the speaker directed microphone. Signals common to both microphones are subtracted out, leaving mostly speech signal, which is then amplified and transmitted. Microphones nowadays may be wireless. Their electrical output is transmitted as an electromagnetic wave to a receiver. Wireless mikes are becoming increasingly popular as their fidelity improves with technological advances. Generally, neither microphones nor ears capture all of the information in a signal when transducing its mechanical vibrations into an electrical signal, as discussed in [Ref. 3].

5. Language And Understanding

Language comprehension by a machine is one of the areas of concern to artificial intelligence (AI) experts. Their opinions range from "yes, it's possible and it's already happening" to "no, it's impossible." The question of whether a computer can be conscious enters into the equation, with respected scholars arguing all sides of the issues. Certainly there are degrees of understanding, putting the question of consciousness aside. It is useful to note the two extremes, where one would not believe that an automobile understands that it's supposed to stop when the brake is applied except, perhaps, metaphorically as discussed in [Ref. 5]. While the other, would believe that a computer capable of passing the Turing test would be said to understand language. The Turing test is conducted as follows: Behind two screens are a computer and a human being. An

interrogator attempts to decide which is which (who is who?) by asking questions and evaluating the answers. The computer is said to have passed the test when the interrogator is unable to do so in a decisive manner. To date, no computer has come close to passing the Turing test, as outlined in [Ref. 6]. There is much argument in the academic world about the legitimacy and even the possibility of such a test, as discussed in [Ref. 3].

Between the two extremes are computers that take as input complex commands in English (and other languages) and respond in complex ways. For example, in a context of data about naval ships, a computer could answer spoken questions such as, "What's the Mercury's average cruising speed?" or "What is the name and c-code of the carrier in the Siberian Sea?" as discussed in [Ref. 7]. The computer answers correctly, however, we may question whether it understood the questions. The computer must recognize most of the words in the question in the sense of being able to repeat them back correctly, much as a shorthand secretary can after taking dictation. The human system of hearing is capable of complex analysis. Through ingenious and highly evolved mechanisms, the ear performs spectral decomposition of auditory input and conveys the information to the brain where it is interpreted. Sounds, such as gunshots, wind rustling in the leaves, telephones ringing, or the allophones of speech are all easily recognized in context. Alone at night in a strange house, the brain may interpret benign sounds as ominous. An acorn rolling across the roof sounds like footsteps in the attic; a loose shutter in the wind is "The Stalker" forcing a window, as discussed in [Ref. 3].

6. Military Applications

Speech recognition systems have been employed by the military in applications ranging from assisting in the repair of tank engines, to accomplishing minor tasks in the cockpit such as adjusting radio frequencies. The cockpit of a modern military aircraft, both fixed and rotary wing (helicopter), is a busy place for the hands and the eyes. Moreover, many of the aircraft systems are too complex to be operated by humans alone, and require the use of computers. The computers, however, are subject to human control and may be instructed through use of a keyboard or touch screen. Manual input strains even further the task load on the hands and eyes. It is a perfect scenario for speech recognition, and indeed, researchers at Wright-Patterson Air Force Base, Fort Ord, Ames Research Center at Moffett Field, and the Aberdeen Proving Ground have been studying how to integrate voice into the command and control needs of the cockpit. Experimental systems have been built and tested for voice-controlled radio tuners, navigation aids, target acquisition systems, and threat-avoidance systems. Under ideal circumstances, voice systems integrate well with other cockpit activity, but conditions in a warplane are never ideal. Pilots may be required to operate their aircraft at high speeds close to the ground, where a wrong decision may lead to catastrophe. They often fly at night and under adverse weather circumstances. The cockpit environment is harsh from the point of view of speech recognition. It is noisy, hot, and full of vibrations. Furthermore, users are under the psychological and physiological factors of stress, fear, and fatigue. Moreover, they may or may not be wearing masks, which affect their voice quality. All of this

conspires to lower speech recognition performance. An avoidance system with voice input that works well in a simulated attack might fail in an actual attack, where the pilot is truly afraid, and the fear causes voice alterations. Advances in microphone technology and increases in the robustness of speech recognition systems have made the use of voice control in the cockpit viable. Nonetheless, one finds such remarks in the literature as "merely adding voice technology to existing displays, or trying to replace visual/motor displays with voice technology on a one-to-one basis can create problems for the pilots." One is reminded of the final scenes of the motion picture Star Wars, where pilot Luke Skywalker eschews his computer-controlled weapon system and stakes the fate of the Galaxy on himself and "The Force." A good reference for the issues of voice in the cockpit is [Ref. 8].

7. Healthcare Applications

The most common deployed application of speech recognition in the healthcare industry is in data entry and report generation. Data entry and report generation using voice recognition in a military hospital was researched and evaluated in [Ref. 9]. Most people who reside in a hospital room are likely to be physically and/or psychologically impaired. Many of the applications of speech recognition in the assistive realm could be used with great benefit in the hospital room. These include television control, bed adjustment, water and ice dispensing, light and door control, voice-activated call button, and so on. Speaker-independent recognition would be required, but the vocabulary could be small and discrete utterances would suffice. The patient could be taught the commands

needed during pre-operation prepping, where patients generally have excessive time on their hands anyway. A surgeon in action is a stereotypic instance of multitasking. Since much surgery today is conducted under a microscope, that instrument must somehow be adjusted when necessary, most likely by the surgeon himself. Some of these functions could be controlled by voice, putting fewer personnel in the operating room, with a concomitant cost saving. The Zeiss Company has experimented with a voice-controlled microscope to be used in ophthalmic surgery, but as of this writing such instruments are rare in practice. "Puff" control microscopes, on the other hand, are commonly found in the operating room. The surgeon blows puffs of air to control the microscope parameters. (Such devices are also in use for severely disabled persons.) They could be considered precursors to the voice-operated microscopes that will undoubtedly become common in the next few years, thanks to the recent advances in speech recognition technology. In a hospital laboratory, or any laboratory where chemicals are handled and sterile conditions are required, technicians find themselves needing "a third hand" to start an exhaust fan, turn on a light, open a door, start or stop a machine, set a timer, and so forth. That third hand could be the vocal cords, as discussed in [Ref. 3].

8. Civilian Applications

In both the United States and the United Kingdom, systems have been implemented that allow travelers to call up a computer, receive travel information, make reservations, and purchase tickets. In the United States, the emphasis is on air travel; in the United Kingdom it is on rail service. The systems are not yet widely used commercially, so they

are midway between hypothetically and commercially successful. The travel information systems involve not only speech recognition on the front end, but a form of artificial intelligence called planning. In essence, the system must have some degree of understanding when you request "information about morning flights from Atlanta to Dallas." Based on that understanding, it plans out the most useful answer it can compute. The system has some leeway as to how to respond. It may decide to give only nonstop flights, or nonstop and direct flights; or it may give all combinations involving only a single stop. Alternatively, it could respond with a question: "First class or coach?" Furthermore, it could ask the traveler for a price limit on the ticket before presenting a choice of flights or ask the travel whether commuter flights should be included, and so on. The system must also be “smart” enough to remember the context of the transaction, so that after answering questions about the flight from Atlanta to Dallas, if the traveler says "What about to Washington?", the system must take this as an inquiry about morning flights' from Atlanta to Washington. One airline is using a voice-driven system for its employees to schedule their flights, as employees are presumably more tolerant and cooperative than the general public. The system was first deployed for corporations and the general public in 1999. In the United Kingdom, a similar spoken language system exists for rail travel, called RailTel. The continuous speech recognizer has a recognition vocabulary of 1,500 words, including 600 station names. The recognizer is adapted to deal with speaker-independent telephone quality speech. Prior to deployment the system was tested by having test subjects interact with the system in a realistic manner. About three-quarters of the calls were successfully completed. As of this writing the system is

still considered experimental. Many banks now permit account information access by telephone, entering data via touch-tones and receiving information via synthetic voice. Speech recognition is desirable where touch-tones are not available, which is 25% of U.S. households and much larger percentages in Europe and Japan, as discussed in [Ref. 3].

9. Future Challenges

The ultimate goal in speech recognition is for the recognition system itself to detect and correct errors since that is what people do. We rarely hear everything said to us perfectly. We are continually applying our human intelligence and knowledge when we recognize speech. The accurate recognition of naturally spoken speech is an unachieved goal and remains the primary aim of speech recognition research. This recognition should be of speech spoken in a typical daily environment such as a busy office and should not require speakers to wear a microphone. It should recognize what most of us do on a daily basis without thinking much about it. When this challenge is met, other even more daunting challenges will appear. The automatic translation of one continuously spoken language into another in approximate real time will appear high on the list.

The ultimate challenge is to recognize multiple speakers speaking simultaneously. At that point, our computers will have exceeded our own abilities. Though you may not be a speech processing professional, you will be able to gauge progress in speech recognition throughout your lifetime by observations in two domains. 1) Communications: Communications companies have invested heavily in speech

recognition. They see the technology both from the point of view of saving labor costs and expanding communications options. One can monitor progress in speech recognition by keeping up with the voice options offered by telephone companies. 2) Personal computer software: At the end of the millennium we find speech recognition becoming a standard option on personal computers and the quality of the software offered follows the state of the art very closely. Most of us have seen a court stenographer at work, either in real life, or on one of the many TV shows or motion pictures that depict courtroom scenes. When a speech recognizer takes over this task, speech recognition truly will have arrived, as discussed in [Ref. 3].

THIS PAGE INTENTIONALLY LEFT BLANK

III. SPEECH RECOGNITION DEVICES

This chapter is designed to provide an overview of speech recognition devices and technologies that could be feasible for operation within a medical first responder's environment researched during the literature review of this thesis to answer the proposed Research Questions in Section 3 of Chapter I.

A. THE VOICE RESPONSE TRANSLATOR

1. History

The National Institute of Justice's (NIJ) Technology Assessment Program Advisory Council (TAPAC) received recommendations in December 1993 from its Weapons and Protective Systems Committee, which identified instant language translation as one of six "immediate" law enforcement technology priorities. As a result of the recommendations, Integrated Wave Technologies, Inc. (IWT) started the development of the Voice Response Translator (VRT). The VRT was designed to be a durable, hands-free device capable of translating voice commands from one language to another. This would allow a police officer to issue commands in English, while the VRT would translate the commands into the native language of the individual, who does not understand English. Models studied included "flip books" used by police officers designed to speak a limited number of phrases in languages such as Spanish. The initial proposal delivered to NIJ stated the device would use approximately 50 phrases. As a result of discussions with the

Oakland Police Department (OPD), IWT expanded the specifications to include about 500 phrases in each of the subject languages. In practice, however, this expansion proved to be cumbersome for initial training with an officer not familiar with the VRT. As a result, the number of phrases was subsequently reduced to about 185. While this number could be increased quite easily from a technical standpoint, later IWT research indicates that the number of languages should be expanded while keeping the number of phrases at this level, as discussed in [Ref. 10].

Police departments have come to rely heavily on telephonic translation services provided by local telecommunications companies. However, the VRT expands the range of initial conversations a police officer can conduct with persons encountered during community policing activities. For example, a lost child can be asked for his or her parents' work numbers or the school of a sibling. Lost children are often found hiding in their homes, so the VRT allows officers to ask for permission to search for them. For victim interviews, the VRT asks whether the perpetrator was a man or woman and to get a specific physical description. Some situations can be resolved by obtaining this type of limited response from persons speaking other languages. But police training should emphasize transition from the VRT to either an in-person translator or the telephonic translation service so that effective communication is maintained and the situation is resolved, as discussed in [Ref. 10].

2. Description

The VRT is the result of six years of research and development led by IWT's Microchip Pioneer, Dr. John Hall , who also designed the first electronic watch, the first computerized heart pacemaker, the first autofocus camera and many other miniaturized electronic devices. The VRT is touted by IWT to achieve performance levels in the areas of speech accuracy, operation in high background noise, miniaturization and low power consumption. The VRT consists of a translator equipped with an external microphone/speaker, a plug-in microphone for pocket use and megaphone that plugs into the translator in place of that microphone. The plug-in microphone replaces the clip-on microphone used with the police version of the translator, as shown in Figure 1.

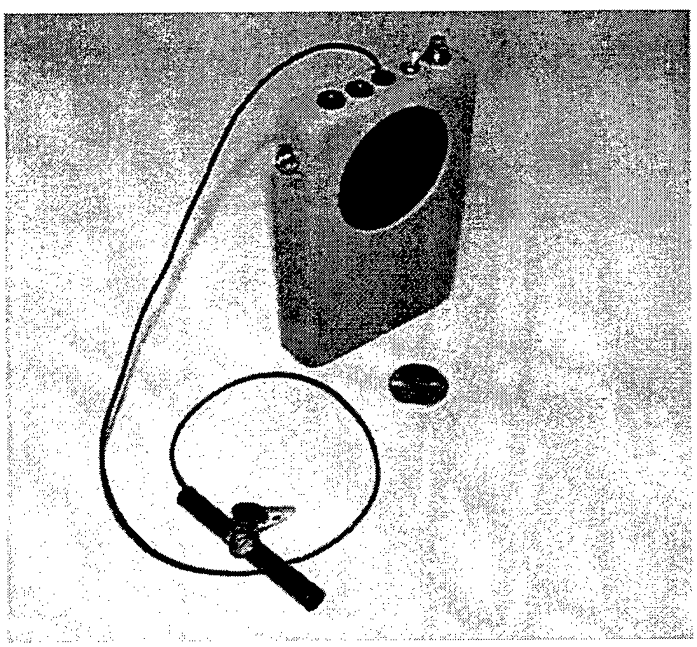


Figure 1. The Voice Response Translator (VRT)

3. Current Status

The current (fourth) generation of the Voice Response Translator has achieved substantial miniaturization. The device is based upon a single-board processing system designed and built by IWT, as shown in Figure 2.

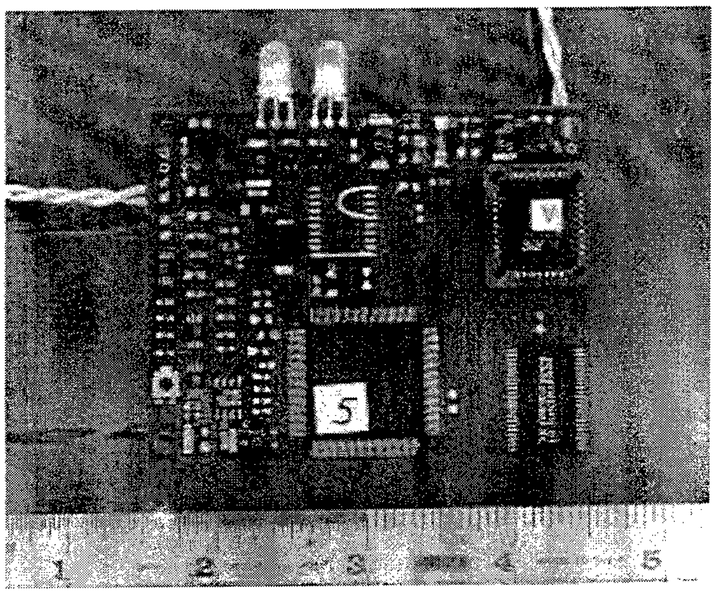


Figure 2. VRT Motherboard (Release approval by IWT)

The functional requirements of the VRT outlined by the Oakland Police Department personnel drove the miniaturization of the device. As a result, the VRT is now able to fit easily within a police officer's shirt pocket, even when space is constricted by the use of a bulletproof vest. Police officers in Oakland stated that the shirt pocket can be viewed as discretionary space where additional equipment such as the VRT can be stored, as belts are already overloaded by bulky equipment, as discussed in [Ref. 10].

B. THE MULTI-LINGUAL INTERVIEW SYSTEM

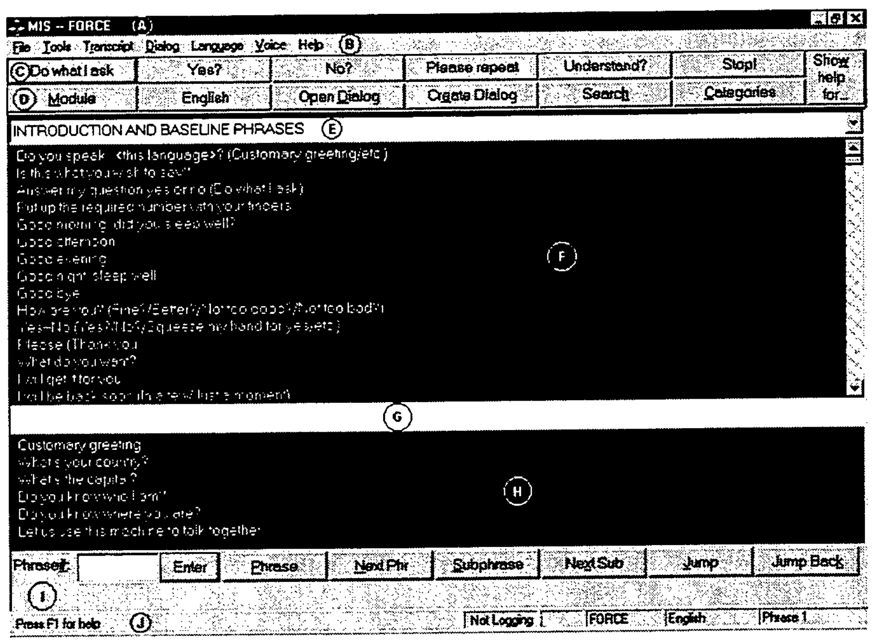
1. History

While stationed in the Persian Gulf during Operation Desert Storm, Captain Lee Morin a United States Navy physician, lacking knowledge of Arabic, expressed the desire to communicate with his non-English speaking patients. Upon returning to the United States he began development of a program that would enable him to communicate with his patients. The program consists of English phrases with corresponding translation in a given language. While stationed at the Naval Operational Medicine Institute (NOMI), Captain Morin was able to get students of foreign nationality to record the necessary phrases based on the NATO translation book for physicians. The first phase of the program was released in 1992. It consisted of a simple point-and-click interface, with three languages available in CD format, called the Medical Language Translator (MLT). By the end of 1995, the MLT was available in 45 languages recorded by native linguists with all phrases organized by medical task on 3 CDs. The Defense Advanced Research Programs Agency (DARPA) became interested in adding voice capability to the device in 1995 and brought together NOMI and Dragon Systems Inc., a commercial speech recognition company based in Newton, Massachusetts, with the intent of providing a speech interface for the MLT. Dragon Systems rewrote the program in *Visual C++*™ for operation under *Windows 95/NT 4.0*™ due to copywriting issues with the MLT, as discussed in [Ref. 11].

The author's first exposure to SR technology was during the 1999 Fleet Battle Experiments (FBE) Echo held in Monterey, California while serving as a member of the Naval Postgraduate School's Assessment Team evaluating the experiments. The author was assigned to assess the effectiveness of the DARPA-One-Way Multi-Lingual Interview System (MIS). During the experiments, the author received the complete history of the MIS from Lieutenant Commanders Eric Rasmussen and Kurt Henry, United States Navy physicians who reported to NOMI and took over the MIS project after Captain Morin. Lieutenant Commander Eric Rasmussen has been the Principal Investigator in Medicine for DARPA over the past five years and the MIS was his first project assignment while serving as the Director of Surface Fleet Medical Programs at NOMI from 1995 to 1997. He was transferred to the Third Fleet as the Fleet Surgeon aboard the Command Ship USS Coronado (AGF-11) located in San Diego, California. The USS Coronado, as part of its mission, evaluates and tests new ideas and concepts, which may be used in future deployment of military strategies and technologies. Lieutenant Commander Kurt Henry was the Special Project Officer for the MIS project from 1997 to 1999. His role in the MIS project and his coordinating efforts for the Language Workshop for the Office of Special Technology under Defense Advanced Research Project Agency (DARPA) led to his current assignment to DARPA as a program manager in the Defense Sciences Office (DSO).

2. Description

The MIS is the second step in a planned approach to minimize problems associated with communication between individuals who do not understand each other's language. MIS is a phrase-based system that plays a pre-recorded wave file (.wav) in the desired language when the desired text file in English is displayed on the computer screen. The .wav file is played by either pointing and clicking on the phrase, a related button with either a mouse or a pen, or optionally by speaking the phrase. Dragon Systems Inc. has developed the voice recognition engine for use in the MIS program. An overview of the main screen layout is displayed below in Figure 3.



A. Caption Bar
B. Menu Bar
C. Hot Buttons
D. Function Buttons
E. Category Line
F. Main Phrases List Box
G. Translation Line
H. Subphrases List Box
I. Operator Buttons
J. Status Bar

Figure 3. MIS Screen Layout

This product has an optional speech interface allowing for hands-free operation and many features of the original MLT were significantly improved and others added. The resulting program was named the MIS and released as the completion of the second phase. Modules for virtually any use can be rapidly developed, and the language files can be produced in-house at little expense. The system can be operated on any size computer from desktop to tablet, thus allowing for great diversification of application, in addition to portability and field use, as discussed in [Ref. 11].

3. MIS In Action

Medical people at the 100th Boston Marathon finish line had some high-tech help from a voice-activated multilingual system similar to that helping U.S. troops in Bosnia. The multi-lingual translator permitted medical workers to use a voice recognition and translation system, loaded into laptop computers, to talk with runners in more than 44 languages. Sets of words, phrases and sentences had been preselected for their utility in medical interviews. "With its large number of foreign entrants, the marathon gave us an opportunity to further test the system in the real world," said John Evans, Hanscom program manager for the Medical Defense Performance Review. The demonstration at the marathon was part of a Transatlantic Telemedicine Initiative led by the Defense Department Medical Defense Performance Review and the Boston-based Atlantic Rim Network, a non-profit information clearinghouse and framework for transatlantic collaboration led by James Barron. "About 50 people were treated using the multi-lingual translators," said Lock Row, senior systems engineer in the MDPR program office. "One

German doctor was extremely enthusiastic about the system as it allowed him to talk easily to foreign patients. He said it was almost like the difference between veterinary and human medicine in that the translator enables the doctor to ask questions such as 'where does it hurt?' and get answers. "Also, the marathon gave our technical people a chance to see how the system works in the real, chaotic world of disaster-like medicine, and therefore they can build systems more responsive to real-world needs," Row said, as discussed in [Ref. 12]. This story is depicted in Figure 4 below.

Figure 4. MIS used in First Aid Station

C. THE VOICE-TO-VOICE TRANSLATION SYSTEM

1. History

The Voice-to-Voice (V-to-V) Translation System of Language Systems Inc. (LSI) is a *SpeechTrans*™ product that was developed to meet the needs of medical, social services, military and law enforcement personnel, and others for rapid, accurate mission-critical translations between English and one or more other languages. The *SpeechTrans*™ products incorporate compact two-way translation software for the Windows 95 or NT environment, for use with notebook or desktop computers. *SpeechTrans*™ can also be configured as a wearable system based on a rugged, belt-mounted computer with a hands-free interface option and other custom features. Depending on the setting in which it is used; the system may require one or two noise-canceling microphones for two-way translation. *SpeechTrans*™ is built around LSI's flexible, customized two-way voice translation engine. It uses speaker-independent continuous speech recognition technology, so that no training is needed and speakers need not pause between spoken words. Instead, each person simply activates the system and speaks naturally to it, as discussed in [Ref. 13].

2. Description

LSI's *SpeechTrans*™ software for law enforcement applications is called *CopTrans*™. A description of *CopTrans*™ functionality is displayed in Figure 5, which shows the initial system display for two-way English-Spanish translation.

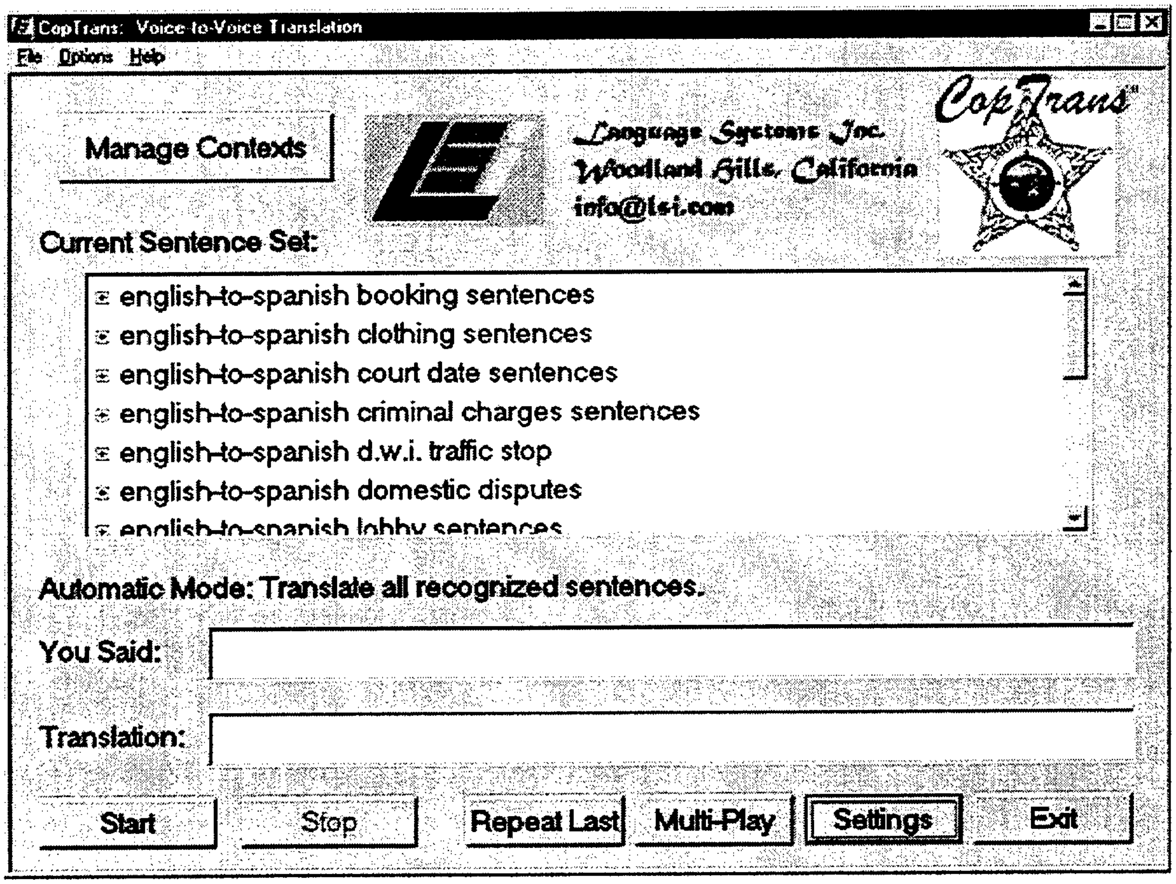


Figure 5. CopTrans Main Screen

The list displayed in Figure 5 represents dialogs, which are appropriate for particular situations. These dialogs are called contexts within the system. To begin using the system, the operator presses the Start button, and opens one or more of these contexts.

3. Operation

The *CopTrans*™ system was designed to recognize multiple users, thus alleviating the requirement for each user to train the unit with their specific voice. LSI's preferred system deployment method is to visit the user's police facility and observe officers in the

actual situations in which they plan to use the system. This allows for system customization to each specific environment, which is followed by on-site training for the officers that will be using the system. The system allows for user modifications to add sentences and phrases that are required, but not included. For example, suppose the user wants to add the new source sentence 'Do you have any contraband?' and the translation '¿Tiene contrabando?' The user will click the *Add* button, which brings up the following display:

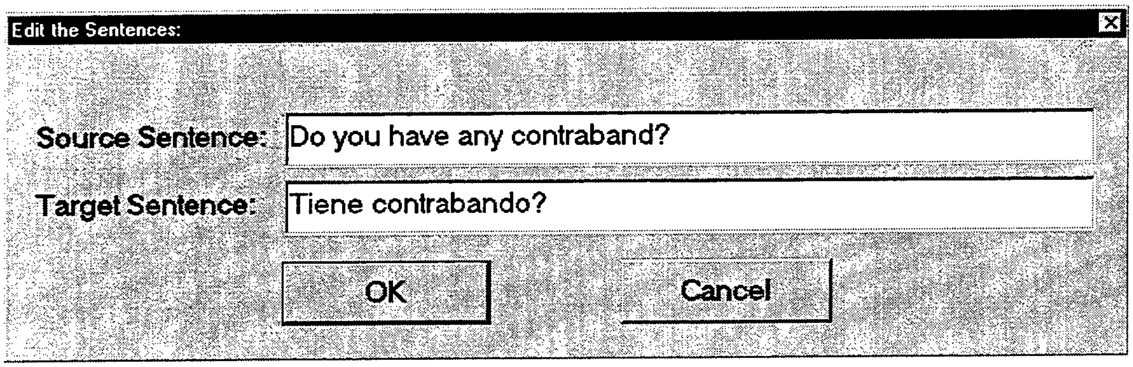


Figure 6. CopTrans Edit Dialog Screen

Note, that the user, as displayed in Figure 6, must supply both source and translation. This version of the system does not do free text translation, so the user must obtain and verify the accuracy of the added translations. By filling in the two windows with the desired sentence pair and clicking on the OK button, the new pair is automatically added to the User context and will be recognized, translated, and spoken just like any other sentence in the system.

CopTrans™ enables two users to converse, each using his own language, as if a human interpreter was present. The system recognizes both what is said and which

language it is said in; it then translates each message into the other language. *CopTrans*™ has the ability to save the spoken input and output in compressed form, like a tape recording, or as a text transcript of the interaction, listing each input utterance as recognized and each output as translated. This allows for later validation, and creates a primary record of the interview, as discussed in [Ref. 13].

D. OTHER SPEECH RECOGNITION DEVICES AND TECHNOLOGY

1. Audio Voice Translator

The Directorate of Combat Developments for the United States Army Chaplain Center and School submitted a requirement to the U.S. Army Combined Arms Support Command to provide military personnel (chaplains, military police, special forces, civil affairs, etc.) with the capability to communicate with indigenous peoples without the use of a human interpreter. The specifics are to develop a speech-to-speech translation capability between English and a range of target languages to support flexible dialogues with allies, host nation military and civilian agencies, indigenous leaders, and civilian populace during the full spectrum of military operations. Also, it must have the capability to verify translation through auditory or visual feedback before executing translation. Translation of both spoken and keyboard input to/from selected host language into both text and audio output. Speaker-independent continuous speech recognition is desired to handle a variety of dialects and voices. Also desired, is adaptation or training period optionally available to improve accuracy in situations of high urgency and low speech

variability. The vocabulary covered by the Audio Voice Translator (AVT) must support dialogues critical in religious support, civil affairs, Special Forces, and military police operations. The AVT must possess the capability for users to add new phrases as needed, and to swap modules depending on language and domains of relevance to the mission. The AVT must be small enough to be hand-held and/or carried in a pocket. All branches will find application for this translation capability vital to operations, especially stability and support operations, internment and resettlement operations, and law and order operations. Additionally, international trade and commerce will likely find a tremendous number of business applications and provide leverage for further technological developments. This research and development program is being performed by Lockheed Martin Federal Systems, Oswego, with subcontractors at Carnegie Mellon University's Language Technology Institute, as discussed in [Ref. 14].

2. Speech Recognition Technology Market Analysis

The speech recognition industry has evolved largely from government-funded research projects in the U.S. and elsewhere. In the U.S., the most well known company has been Dragon Systems, Inc. acquieved in March 2000 by the Lernout & Hauspie (L&H). L&H has also acquired the U.S. speech recognition company Kurzweil. By way of these acquisitions, as well as its own research and business development efforts, L&H could be considered the leader in the speech recognition industry. Despite this investment, L&H has been unable to produce either call center systems or hand-held recognition technology.

IBM, Unisys, Microsoft and Apple have also devoted significant resources to developing and marketing speech recognition products. Microsoft, in addition to its research efforts, acquired Entropic Research Laboratory, Inc., a speech recognition technology development firm in mid 1999.

Automobile companies, call center companies and cellular telephone handset manufacturers have made similar commitments to develop speech recognition as ancillary features to their main product lines. Lucent in early 1999 announced a new unit, Lucent Speech Solutions, to focus on speech products in communications networks. Philips, a large Dutch electronics company, offers speech recognition for Windows-based applications through its Austria-based subsidiary, Philips Speech Processing.

Recent important business developments in the Interactive Voice Response (IVR) area include:

a. Nuance Communications which announced it has integrated its family of speech recognition and speaker verification products with Lucent Technologies' CentreVu Response Solutions suite of offerings. The companies will jointly market those IVR products, through value-added resellers (VARs), to call centers around the world. This alliance will give Lucent customers the option to choose natural language speech technology from Nuance or from Lucent Speech Solutions.

b. Omnitel and Philips which announced Omnitel 2000, an Internet and Communications Service Provider in Italy. The two companies bring together mobile

telecommunications, speech recognition technology and the Internet in the platform, which is available to all Omnitel Pronto Italia mobile customers and customers of other Italian telecom operators.

c. Lucent Technologies, which recently created a new unit, Lucent Speech Solutions, to focus on speech products in communications networks. The new unit will deliver speech recognition, personal agent technology, and text-to-speech synthesis for a wide range of customer applications, all based on Bell Labs speech technology.

d. Unisys Corporation and Microsoft Corporation which have announced a marketing and technology alliance that promises to broaden the market for advanced desktop and telephony speech applications. These two companies are working together to accelerate the adoption of the Speech Application Programming Interface (SAPI) for speech applications, providing software developers with tools and support programs that will make speech-based technology easier to deploy. Unisys has created a Natural Language Understanding (NLU) Services organization that will focus specifically on consulting and application development for customer interaction.

e. Nuance and Unisys which announced a broad technology and services agreement to deliver complete, high-quality speech recognition solutions to call centers and communication service providers.

Similar companies have shown interest in the development of hand-held speech recognition, but technical challenges relating to background noise immunity and accuracy

have limited this market to marginal product features. With great fanfare, some large companies with investments in speech recognition and mobile technologies announced in 1999 the formation of the Voice Technology Initiative for Mobile Enterprise Solutions (VoiceTIMES). VoiceTIMES' stated goal is to coordinate the technical requirements needed for companies to build and deploy solutions using voice technologies and hand-held mobile devices. Inaugural VoiceTIMES alliance members include Dictaphone, e.Digital, IBM, Intel, Norcom Electronics, Olympus and Philips, as outlined in [Ref. 15].

THIS PAGE INTENTIONALLY LEFT BLANK

IV. DEMONSTRATION AND EVALUATION

This scenario will begin with a first responder arriving on the scene after receiving a humanitarian call for help from the nearest United Nations support station who was recently attacked by hostile rebels. Once on the scene, the first responder will identify himself as an emergency response team member ready to assist the foreign casualties while communicating in his/her native language.

The demonstration and evaluation of the VRT was conducted at the Defense Language Institute from October 25, 2000 to November 2, 2000. It consisted of an evaluation of the prerecorded languages of Spanish, Vietnamese, and Loa to verify the correctness of the statements and their content. MSgt Jose Sanchez is a Military Language Instructor at the DLI, provided the evaluation of the prerecorded Spanish used with the VRT. MSgt Kelly Ray is a Military Language Instructor at the DLI, provided the evaluation of the prerecorded Vietnamese used with the VRT. The demonstration began with a brief explanation of the concepts of the VRT and the approach to be used for this evaluation.

A demonstration and evaluation survey tool was developed to assist in evaluating the feasibility of technologies being researched and developed such as the VRT to be used in the operating environment of the medical first responder. This demonstration and evaluation was developed to answer the Research Question "What SR technologies are available for operating within a medical first responder's environment?" The environment of the medical first responder, for the purposes of this study, is one where

the medical personnel are assigned to a unit such as the Fleet Marine Forces. In this environment, the first responder is responsible for maintaining the medical supplies needed to treat his Marines, which leads to little or no additional cargo space for extra supplies or equipment. The research and development efforts recommending technologies for improving the performance and abilities of the first responder in carrying out his/her mission should always carefully consider the limitations of the environment. This environment also consists of the first responder communicating his/her ability to render first aid to a non English-speaking patient. This scenario arises very often when responding to a humanitarian, multi-national, overseas operation or exercise where the medical personnel are tasked with treating and supporting the needs of patients other than the United States Armed Forces.

To ensure that the evaluation of the VRT would be realistic in its approach of considering the needs of the first responder, the following criteria was used:

- Limited amount of time was devoted to user training,
- No prior speech recognition or computer knowledge required,
- The device must be portable and lightweight,
- The device must be durable for a field environment.

The VRT was given to Lieutenant John Kendrick, the Officer-in-Charge of the Navy Medical Administrative Unit of the Presidio of Monterey Medical Clinic. Next, the VRT was given to four corpsmen and two medics with varying operational experience as medical first responders. Each corpsman/medic was instructed to evaluate the VRT for its usefulness as an assistive device during an initial patient assessment tool when evaluating

a non-English speaking patient in a field environment. The purpose of this evaluation was intended to ascertain the corpsmen ability to self-train on the VRT unit by utilizing the instruction manual without the assistance of a human instructor. This element was used because the intended device for use should be easy to use and require minimum training time. This is similar to the most realistic approaches used for deploying such a unit because training is always limited due to all other required training imposed on the first responder.

THIS PAGE INTENTIONALLY LEFT BLANK

V. DEMONSTRATION AND EVALUATION RESULTS

This chapter presents the findings from the demonstration and evaluation of the Voice Response Translator (VRT) held at the Defense Language Institute Foreign Language Center and the Navy Medical Administrative Unit located on the Presidio of Monterey Annex. The Perception Questionnaire was the data instrument used to evaluate the viability, perception and performance of the VRT. Section A covers the data instrument and collection procedures. Section B covers the findings of the questionnaire.

A. PERCEPTION QUESTIONNAIRE

1. Instrument Development

A perception questionnaire was developed to assess the medical first responder's (Navy Corpsmen) feasibility of the VRT being used in a medical field environment. The data gathered from this questionnaire addresses the following research question: What other SR assisted technology being used in other than a medical environment could be feasible for operation within a medical first responder's environment? The perception questionnaire was made up of four sections designed to solicit a general summary opinion from the participants who provided a response that most closely corresponded to their opinion on the questions presented by a scoring scale ranging from strongly disagree (1) through strongly agree (5). An example of the perception questionnaire is provided in Appendix A.

2. Collection Procedures

The questionnaire was distributed to two foreign language staff members and eight medical personnel located on the Presidio of Monterey Annex during the evaluation period from October 20th through November 5th 2000. They were told that the questionnaire was collecting data on the feasibility of SR devices for thesis research at the Naval Postgraduate School, Monterey, California. To ensure that the device would be evaluated in a typical pre-deployment scenario, The VRT was issued with a brief training manual and the participants were instructed to review the manual, train and use the device, and provide their overall opinion of the VRT.

B. FINDINGS

Distributing the questionnaire to a larger medical community was impossible due to the resources required and time constraints. Therefore, these findings are based on the small sample size of Navy Corpsmen available at the Navy Medical Administrative Unit. Also, two foreign language staff members from the Defense Language Institute Foreign Language Center were used to provide their opinion on the pre-recorded translated statements for correctness and accuracy. There were four distinct phases where each participant had to circle a number that most closely corresponded to his or her opinion about the question being asked. The scoring scale described below is Strongly Disagree =1; Disagree = 2; Neutral =3; Agree =4; and Strongly Agree =5. The findings of those four phases are described below.

1. Knowledge Phase

The knowledge phase was developed to ascertain the prior knowledge of the participants concerning computer speech technology, foreign languages and translators. The results of the knowledge phase as described in Table 2.

	Score Totals				
Questions	1	2	3	4	5
I am familiar with computer speech technology	3	2	3	2	0
I am familiar with foreign language translators	4	4	2	0	0
I speak a foreign language	8	0	0	0	2

Table 2. Knowledge Questions

Only 20% of the respondents were familiar with computer speech technology and none were familiar with foreign language translators. Eighty percent of the respondents did not speak a foreign language; this 80% represented the medical personnel participating in the evaluation, which is the targeted audience for this study. Two Military Language Instructors from the Defense Language Institute Foreign Language Center represented the 20% of the respondents with foreign language skills. The languages evaluated were Spanish and Vietnamese.

2. Training Phase

The training phase was developed to ascertain the opinion of the participant concerning the training instructions and the VRT's training process. The results of the training phase are described in Table 3.

Questions	Score Totals 1	2	3	4	5
The instructions for training the VRT were easy to follow	0	0	2	5	3
The VRT had no problems recognizing my voice commands	3	5	1	1	0
Once recording began, the training evolution took about 10 minutes	0	2	1	5	2
Training the VRT was an easy process	0	1	4	4	1

Table 3. Training Questions

Eighty percent of the respondents said that the instructions for training the VRT were easy to follow, but of that 80%, only 12.5% said that the VRT performed according to its instructions. Eighty percent admitted that the VRT had no problems recognizing their voice commands. Seventy percent of the respondents said that once the recording began, the training evolution took about 10 minutes. Fifty percent of the respondents admitted that training the VRT was an easy process.

3. Operational Phase

The operational phase was developed to ascertain the opinion of the participants concerning the overall performance of the VRT. The results of the operational phase are described in Table 4.

Questions	Score Totals 1	2	3	4	5
The VRT had no problems recognizing my voice commands	6	3	1	0	0
The VRT had no problems switching from one language to another	2	1	5	1	1
Voice commands had to be repeated often	0	0	0	3	7
The translated statements sounded clear during operation	0	0	1	6	3
Translated statements were prerecorded correctly	0	0	2	5	3
The VRT was easy to use and operate	0	2	4	3	1

Table 4. Operational Questions

One hundred percent of the respondents said that the VRT had problems recognizing their voice commands issued to the VRT and that they had to repeat their voice commands several times. Ninety percent of the respondents admitted that the translated statements sounded clear during operation. Eighty percent admitted that the VRT

translated statements were prerecorded correctly, however, only 40% said the VRT was easy to use and operate.

4. Evaluation Phase

The evaluation phase was developed to ascertain the opinion of the participant concerning the feasibility of VRT operating within a Medical First Responder's Environment. The results of the evaluation phase are described in Table 5.

	Score Totals				
Questions	1	2	3	4	5
The VRT performed as intended according to its instructions	0	3	6	1	0
The VRT is a lightweight portable device	0	0	0	2	8
The concept of a language translation assisted device is a good idea	0	0	0	2	8
I can envision the VRT being useful in a foreign language environment	0	0	1	1	8

Table 5. Evaluation Questions

Only 10% said that the VRT performed as intended according to its instructions, while 100% of the respondents said that the VRT is a lightweight portable device and that the concept of a language translation assisted device is a good idea. Finally, 90% of the respondents admitted that they could envision the VRT being useful in a foreign language environment. All of the respondents said that devices such as the VRT are needed and a great idea. However, they also emphasized that further work is needed on the VRT's ability to recognize speech. Follow-up conversations with the respondents revealed that most of them did not read the training manual in its entirety and there were no attempts to repeat the initial training of the VRT if their voice was not being recognized. This scenario is a very good representation of exactly how most users (such as the first responder) would use the device in a deployment situation. There is always a limited

amount of time available for additional training above and beyond the required predeployment training.

5. Conclusion

This evaluation revealed that planned training procedures for the VRT might not be adequate to obtain the users initial voice template. Most potential users were reluctant to devote time reading the entire training manual that recommends to users to re-train the Unit for better performance and as a result didn't retrain the Unit for better performance. The lack of proper training resulted in significantly degraded performance. Based upon this experience, this study recommends a five-minute training video or other training aids that replace or complement the written manual.

VI. SUMMARY AND RECOMMENDATIONS

A. SUMMARY OF FINDINGS

In this section, the findings from Chapters III through V will be used to answer the research questions proposed in this thesis.

What are the SR technologies available for operating within a medical first responder's environment? There are many ongoing research efforts in SR technology that could be used in a field medical environment. However, the VRT was the only miniaturized device discovered through this research that was evaluated to be feasible for operating in the field medical environment, as discussed in Chapter IV of this study.

What are the SR technologies currently being used in a medical first responder's environment? The MIS is the device that is currently being used in humanitarian, shipboard and other medical operating environments.

What SR are the technologies available for operating within a medical first responder's foreign language environment? The VRT was evaluated in detail because it was the only SR device available through this research that is miniaturized, durable and capable of operating in a field environment.

What are the SR technologies currently being used in a medical first responder's foreign language environment? The MIS is the device that is currently being used in humanitarian, shipboard and other medical operating environments that has foreign language capabilities.

What other SR assisted technologies could be used for operation within a medical first responder's environment? All devices discussed in Chapter III of this study are feasible for operation within a medical environment, but the VRT is the only device researched that was practical for operating in a medical first responder's environment.

B. RECOMMENDATION FOR NAVY MEDICINE

The Navy Medical Department should be involved in researching SR technologies available for assisting the medical first in a foreign language environment. A prudent research approach for Navy Medicine when exploring SR technologies for use in a foreign language environment is to include other military support functions having similar requirements, such as Chaplains, Supply Corps, and Military Police. Combining research and development efforts will ensure that solutions found meet specific hardware and software requirements. In addition, using task specific domains will alleviate over tasking the device, which usually occurs when the mission requirements for the device are not clearly defined. Finally, it is very important to recognize the importance of the Defense Language Institute Foreign Language Center, when researching foreign language SR technologies because it can provide experts needed for product evaluation. The best starting point for research in SR technologies is DARPA, which leads the Department of Defense efforts in research and development.

APPENDIX A. PERCEPTION QUESTIONNAIRE

THE VOICE RESPONSE TRANSLATOR

Perception Questionnaire
Prepared by LT Leroy W. Harris Jr.
Naval Postgraduate School
Thesis Research

This perception questionnaire is provided to ascertain your opinion of the Voice Response Translator (VRT) as a part of my Thesis Research pertaining to a "Feasibility Study of Speech Recognition Devices for operating within a Medical First Responder's Environment."

[PLEASE CIRCLE THE NUMBER THAT MOST CLOSELY CORRESPONDS TO YOUR OPINION]
[Strongly Disagree –1, Disagree – 2, Neutral –3, Agree –4, Strongly Agree –5]

KNOWLEDGE PHASE

I am familiar with computer speech technology?	1	2	3	4	5
I am familiar with foreign language translators?	1	2	3	4	5
I speak a foreign language?	1	2	3	4	5

TRAINING PHASE

The instructions for training the VRT were easy to follow?	1	2	3	4	5
The VRT had no problems recognizing my voice commands?	1	2	3	4	5
Once recording began, the training evolution took about 10 minutes?	1	2	3	4	5
Training the VRT was an easy process?	1	2	3	4	5

OPERATIONAL PHASE

The VRT had no problems recognizing my voice commands?	1	2	3	4	5
The VRT had no problems switching from one language to another?	1	2	3	4	5
Voice commands had to be repeated often?	1	2	3	4	5
The translated statements sounded clear during operation?	1	2	3	4	5
Translated statements were prerecorded correctly?	1	2	3	4	5
The VRT was easy to use and operate?	1	2	3	4	5

EVALUATION PHASE

The VRT performed as intended according to its instructions?	1	2	3	4	5

The VRT is a lightweight portable device?	1	2	3	4	5
The concept of a language translation assisted device is a good idea?	1	2	3	4	5
I can envision the VRT being useful in a foreign language environment?	1	2	3	4	5

THANK YOU FOR YOUR PARTICIPATION

APPENDIX B. SUMMARY COMMENTS

27 Nov 00

From: Officer in Charge, Naval Medical Administrative Unit, Monterey, CA 93944
To: Leroy Harris, LT, MSC, USN, Naval Postgraduate School

Subj: VOICE TRANSLATOR SYSTEM

1. Per your request, Naval Hospital Corpsmen and Army Medics tested the Voice Translator System. They reviewed it for ease of use, quality of translation and overall usefulness in a medical triaging system. Their summarized comments are as follows:

LT John Kendrick, MSC, USN – the concept is great and I feel it should be pursued vigorously but I did experience problems with voice recognition. Although I am not a medical provider, I feel any delays with a system such as this could be problematic. My recommendation is to correct the problems and implement.

HM2 Thomas Luttrell, USN – the system would be extremely valuable in a field triage environment, especially during an emergency situation. The system I reviewed showed great promise but I had difficulty with effective translation and systematic use. I feel this system should be pursued but the kinks need to be worked out.

HM2 (FMF) Jason Ivie, USN – it is a great idea and should be used but the system has a few problems. It had a difficult time picking up my voice and the translation. This system would be a great use in a foreign country if it works properly.

HM2 (FMF) Cory Whittle, USN – the machine is a great idea and would serve any corpsmen well in a foreign country. I had trouble with it recognizing my voice and the directions were sometimes confusing. It should be pursued but the kinks need to be worked out.

HM3 (FMF) Jason Tetzlaff, USN – the system will be a great thing once the problems with voice recognition are worked out. I spent too much time trying to say things that in an emergency situation, I wouldn't have time for.

SPC Nicholas Starkey, USA – the system is a good thing and will be of great value once the voice recognition process is fixed. I spent too much time trying to get it to recognize my voice. Under battlefield conditions, I wouldn't have time to repeat myself.

SPC John Gary, USA – the system is a great tool but the problems with voice recognition need to be fixed prior to battlefield implementation. I had difficulty with it recognizing my voice and it got frustrating after a while. Its potential is unlimited.

2. If you need additional information, please contact me at (831) 242-7542; DSN 878 or via e-mail at jpkendri@nps.navy.mil.

J. P. KENDRICK

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

1. Navy Medicine Strategic Plan, http://bumed.med.navy.mil/meddept.pdf, dated 12 October 2000.

2. Klatt, D., "Review of the ARPA Speech Understanding Project," Journal of the Acoustic Society of America, Volume 62, (1977): 1345-1366.

3. Rodman, Robert D., "Computer Speech Technology," ARTECH House, Inc., Norwood, MA, 1999.

4. Winograd, T., "Language as a Cognitive Process," Addison-Wesley Publishing Company, Reading, MA, 1983.

5. Crevier, D., "AI: The Tumultuous History of the Search for Artificial Intelligence," New York, NY, 1993.

6. Turing, A. M., "Computing Machinery and Intelligence," Mind, Volume 59, 1950.

7. Lee, K. F., "Automatic Speech Recognition," Kluwer Academic Publishers, Norwood, MA, 1989.

8. Voorhees, J. W., and Bucher, N. M., "The Integration of Voice and Visual Displays for Aviation Systems," Journal of the American Voice Input/Output Society, Vol. 1, No. 1, June, 1984.

9. Threet, E., "Economic Evaluation of Voice Recognition (VR) for the Clinician's Desktop at the Naval Hospital Roosevelt Roads (NHRR)," Masters Thesis, Naval Postgraduate school, Monterey, California September, 1997.

10. Integrated Wave Technologies, http://www.i-w-t.com, dated 25 October 2000.

11. Naval Operational Medicine Institute, http://www.namrl.navy.mil, dated 11 October 2000.

12. Hanscom Air Force Base, http://www.hanscom.af.mil/esc-pa/news/1996/ apr96/ boston.htm, dated 12 October 2000.

13. Language System Incorporated, http://www.lsi.com, dated 13 October 2000.

14. U.S. Army Combined Arms Support Command, http://www.cascom.army.mil /cssbl, dated 20 November 2000.

15. McCune, Timothy, "Market Analysis", Eagan, McAllister Associates, June 2000.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center 2
   8725 John J. Kingman Road, Ste 0944
   Fort Belvoir, VA 22060-6218

2. Dudley Knox Library 2
   Naval Postgraduate School
   411 Dyer Road
   Monterey, California 93943-5101

3. Professor Monique P. Fargues 2
   ECE Department, Code EC/Fa
   Naval Postgraduate School
   Monterey, CA 93943

4. Dr. Ray T. Clifford 2
   Defense Language Institute
   Presidio of Monterey, CA 93944-5006

5. LT Leroy W. Harris 3
   CINCPACFLT, Code N01M3
   250 Makalapa Drive
   Pearl Harbor, HI 96860-3131

6. Professor Douglas E. Brinkley 1
   Systems Management Department, Code SM/Bi
   Naval Postgraduate School
   Monterey, CA 93943

7. Commanding Officer 1
   Naval Medical Information Management Center
   8901 Wisconsin Ave. Bld. 27
   Bethesda, MD 93943-5101

8. Mr. Tim McCune 1
   Eagan, McAllister Associates, Inc.
   1500 N. Beauregard Street
   Alexandria, VA 22311

9. Bureau of Medicine and Surgery 1
   Attn: MED-08
   2300 E. Street, NW
   Washington, DC 20372-5300

10. Chairman 1
ECE Department, Code EC
Naval Postgraduate School
Monterey, CA 93943

11. Chairman 1
Information Systems Academic Group, Code IS
Naval Postgraduate School
Monterey, CA 93943